BMT ALL Exams 2021/2022 - Solution

Basic Mathematical Tools For Imaging and Visualisation (Technische Universität München)

Computational Imaging and Inverse Problems
Department of Informatics
Technical University of Munich

Esolution
Place student sticker here

Note:

- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Basic Math Tools for Imaging and Visualization

Exam:	IN2124 / Homework-1-LA	**Date:**	Wednesday 24th November, 2021
Examiner:	PD Dr. Tobias Lasser	**Time:**	09:00 – 09:00

Working instructions

- Make sure to **sign your name** at the top of this page.
- This graded homework assignment consists of **8 pages** with a total of **5 problems**.
- The total amount of achievable credits in this graded homework assignment is 40 credits.
- Use only the pages provided in this PDF. Do not insert additional pages. There are two pages at the end with additional space for solutions. If you use them, make sure to mark which (sub)problem your answers are related to, and note in the original problem that your solution continues elsewhere.
- Allowed resources:
  - all materials, for example books, scripts, slides, sources from the internet, programs written by yourself
  - **not** allowed is help from any third person in any form
  - **not** allowed is plagiarism
- Subproblems marked by * can be solved without results of previous subproblems.
- **Answers are only accepted if the solution approach is documented.** Give a reason for each answer unless explicitly stated otherwise in the respective subproblem.
- Do not write with red or green colors nor use pencils.

Problem 1 Eigenvalues (9 credits)

0 1 2 3

a)* Let $A \in \mathbb{R}^{n \times n}$ for some $n \in \mathbb{N}$. Write down the **exact** definition of what an eigenvalue of A is.

If there exists $x \in \mathbb{C}^n$ with $x \neq 0$such that

$$Ax = \lambda x$$

for some $\lambda \in \mathbb{C}$, then λ is called an eigenvalue of A.

0 1 2 3 4 5 6

b)* Consider the matrix

$$B = \begin{pmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

Compute all eigenvalues of B and their corresponding eigenspaces. Determine the algebraic and geometric multiplicities of all eigenvalues. Decide if B is a defective matrix. Include **all relevant** computational steps in your answer and justify your answers.

The characteristic polynomial $\det(\lambda I - B)$ evaluates to

$$\det \begin{pmatrix} \lambda - 4 & 0 & -1 \\ 2 & \lambda - 1 & 0 \\ 2 & 0 & \lambda - 1 \end{pmatrix} = (\lambda - 4)(\lambda - 1)(\lambda - 1) + 2(\lambda - 1) = \lambda^3 - 6\lambda^2 + 11\lambda - 6$$
$$= (\lambda - 1)(\lambda - 2)(\lambda - 3),$$

The eigenvalues of B are thus $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$, all with algebraic multiplicity 1.

Eigenspace of $\lambda_1 = 1$: we solve $(\lambda_1 I - B)x = 0$ for $x = (x_1, x_2, x_3)^T \neq 0$, i.e.

$$\begin{pmatrix} -3 & 0 & -1 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The second column is all zeros, hence x_2 is free, we have $2x_1 = 0$ and $-3x_1 - x_3 = 0$. Hence, the eigenspace of λ_1 is thus spanned by $(0, 1, 0)^T$ with geometric multiplicity 1.

Eigenspace of $\lambda_2 = 2$: we solve $(\lambda_2 I - B)x = 0$ for $x = (x_1, x_2, x_3)^T \neq 0$, i.e.

$$\begin{pmatrix} -2 & 0 & -1 \\ 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The third row is a multiple of the first row, so we ignore it. We have $-2x_1 - x_3 = 0$ and $2x_1 + x_2 = 0$, hence the eigenspace of λ_2 is spanned by $(1, -2, -2)^T$ with geometric multiplicity 1.

Eigenspace of $\lambda_3 = 3$: we solve $(\lambda_3 I - B)x = 0$ for $x = (x_1, x_2, x_3)^T \neq 0$, i.e.

$$\begin{pmatrix} -1 & 0 & -1 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The third row is again a multiple of the first row. We have $-x_1 - x_3 = 0$ and $2x_1 + 2x_2 = 0$, hence the eigenspace of λ_3 is spanned by $(1, -1, -1)^T$ with geometric multiplicity 1.

Geometric and algebraic multiplicity are identical for all eigenvalues, hence the matrix B is not defective.

Problem 2 Matrices (10 credits)

a)* Consider the matrix

$$C = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for $\theta \in [0, \pi]$. Compute all **real** eigenvalues of C. Explain what this result means in a geometric sense.

0 1 2 3 4

The characteristic polynomial $\det(\lambda I - C)$ evaluates to

$$\det \begin{pmatrix} \lambda - \cos\theta & \sin\theta \\ -\sin\theta & \lambda - \cos\theta \end{pmatrix} = \lambda^2 - 2\lambda\cos\theta + \cos^2\theta + \sin^2\theta = \lambda^2 - 2\lambda\cos\theta + 1,$$

hence the eigenvalues of C are $\lambda = \cos\theta \pm i\sin\theta$.
That means the eigenvalues are only real for $\sin\theta = 0$, or $\theta \in \{0, \pi\}$. We have $\lambda = 1$ for $\theta = 0$, and $\lambda = -1$ for $\theta = \pi$.

C is a two-dimensional rotation matrix. Geometrically this means:

- for $\theta \in (0, \pi)$: no non-zero vector is mapped onto a scalar multiple of it (which is obvious, as C is a rotation by θ).
- for $\theta \in \{0, \pi\}$, the matrix C is the identity / negative identity, and hence has eigenvalues ± 1.

b)* Consider the matrix $D \in \mathbb{R}^{n \times n}$ for some $n \in \mathbb{N}$. Assume further, that there is an orthogonal matrix $P \in \mathbb{R}^{n \times n}$ that diagonalizes D. Show that D has to be symmetric.

0 1 2 3

P diagonalizes D, hence we have

$$\Sigma = P^{-1}DP,$$

for some diagonal matrix $\Sigma \in \mathbb{R}^{n \times n}$.
We know that $P^{-1} = P^T$, since P orthogonal, hence $\Sigma = P^T DP$ and thus $D = P\Sigma P^T$.
Finally, we have

$$D^T = \left(P\Sigma P^T\right)^T = P\Sigma P^T = D,$$

in other words, D is symmetric.

c)* Let $E \in \mathbb{R}^{n \times n}$ for some $n \in \mathbb{N}$ be a diagonalizable matrix. Describe a way to compute E^{1023} using significantly less than 1022 matrix-matrix multiplications.

0 1 2 3

Since E is diagonalizable, there is an invertible matrix P with $P^{-1}EP = D$, where D is diagonal with $D = \operatorname{diag}(d_1, \dots, d_n)$.We have $E = PDP^{-1}$, and hence

$$E^{1023} = \left(PDP^{-1}\right)^{1023} = PDP^{-1} \cdot PDP^{-1} \cdots PDP^{-1} = PD^{1023}P^{-1},$$

where $D^{1023} = \operatorname{diag}(d_1^{1023}, \dots, d_n^{1023})$ is easy to compute.

Problem 3 X-ray computed tomography (8 credits)

Due to a certain pandemic, you are tasked with diagnosing lung diseases in patients using X-ray computed tomography (in short: X-ray CT). The resulting three-dimensional image of the X-ray absorption coefficients will allow your medical partners to do the diagnosis.

0 1 2

a)* You represent the lung area of the patient using a set $V \subset \mathbb{R}^3$, hence the lung absorption coefficients can be represented as a continuous function $f : V \to \mathbb{R}$. For your computer algorithms, this function f has to be discretized somehow into a discrete $\hat{f}$. As an additional difficulty, your X-ray CT scanner has known **non-uniform** scanning characteristics, resulting in higher resolution in specific image areas, while the other image areas have lower resolution.

Formulate a suitable mathematical representation of the discrete $\hat{f}$ based on your knowledge from the BMT course. Explain and justify your choice.

Since the resolution of the scanner is variable, a representation using uniform voxels is not suitable. Instead, we use $N \in \mathbb{N}$ non-uniform voxels, represented by voxel functions $b_i : V \to \mathbb{R}$ that are 1 inside the i-th voxel and 0 outside, for $i = 1, \ldots, N$. Each b_i is chosen such that it matches the scanner properties closely.

Our discretized image is then

$$\hat{f}(z) = \sum_{i=0}^{N} x_i b_i(z) \qquad \text{for } z \in V,$$

where the coefficients (x_i) denote the X-ray absorption coefficient of the i-th voxel b_i.

0 1 2 3

b) A partner from physics gives you a linear model $\mathcal{M}_j : f \mapsto m_j$ that explains how the X-ray CT scanner maps the lung area f to a detector measurement m_j. The X-ray CT scan takes $M \in \mathbb{N}$ of these measurements, i.e. $j = 1, \ldots, M$. Explain in detail how the series expansion approach from the BMT course takes all these ingredients (i.e. your image representation from a), the model $\mathcal{M}_j$, and the measurements $m = (m_j)$) and produces a linear system $Ax = m$.

We have $f \approx \hat{f} = \sum_{i=1}^{N} x_i b_i$ from a). Applying $\mathcal{M}_j$ to $\hat{f}$ yields

$$m_j = \mathcal{M}_j(\hat{f}) = \mathcal{M}_j\left(\sum_{i=1}^{N} x_i b_i\right) = \sum_{i=1}^{N} x_i \mathcal{M}_j(b_i),$$

where we used the linearity of $\mathcal{M}_j$. Using $a_{ji} := \mathcal{M}_j(b_i)$and $A := (a_{ji})$ we receive the linear system

$$Ax = m,$$

where $x = (x_i)$ and $m = (m_j)$.

0 1 2 3

c) The detector in your X-ray CT scanner has 1000×1000 pixels, and it measures from 2000 different angles covering the full circle around the patient. Furthermore, you choose to use 10^9 discretized absorption coefficients in a) to represent the lung area. What is the matrix size of the so-called *system matrix* A from the linear system $Ax = m$ in b)? What implication does this have for a practical implementation on a computer?

There are $1000 \times 1000 = 10^6$ detector pixels that are measured $2000 = 2 \cdot 10^3$ times. Hence the system matrix A has $2 \cdot 10^9$ rows. The stated 10^9 is the number of absorption coefficients, i.e. the size of $x = (x_i)$, and hence the number of columns of A. In total, $A \in \mathbb{R}^{2 \cdot 10^9 \times 10^9}$.

In other words, A has $2 \cdot 10^{18}$ entries, which for 32bit floats requires roughly 7 exabyte of memory. This means that A does not fit into any computer's memory!

Problem 4 Solving inverse problems (9 credits)

We continue the lung imaging with X-ray CT scenario from problem 3. We use the linear system $Ax = m$ as the forward model, where $x = (x_i) \in \mathbb{R}^N$ denotes the absorption coefficients, $m = (m_j) \in \mathbb{R}^M$ denotes the detector measurements, and $A \in \mathbb{R}^{M \times N}$ the system matrix, with $N, M \in \mathbb{N}$ and $2N > M > N$.

a)* Assume you have the singular value decomposition (SVD) of $A = U\Sigma V^T$, where $U = (u_1, \dots, u_n) \in \mathbb{R}^{M \times N}$ with orthogonal columns, $V = (v_1, \dots, v_n) \in \mathbb{R}^{N \times N}$ orthogonal, and $\Sigma \in \mathbb{R}^{N \times N}$ diagonal with the singular values σ_i on the diagonal. Assume furthermore you have an instance of noisy detector measurements $m = (m_j) \in \mathbb{R}^M$.
Write down the solution x_{SVD} of $Ax = m$ using the SVD. Describe briefly how the resulting image x_{SVD} will look like. Use mathematical terminology and your formula for x_{SVD} to explain why this happens. State a technique to improve the results, and explain why it helps. Keep your explanations brief and to the point!

0 1 2 3 4

We know that $x_{SVD} = \sum_{i=1}^{N} \frac{u_i^T m}{\sigma_i} v_i$ is the SVD solution of the linear system.

A is ill-conditioned, as always in computed tomography, meaning $\text{cond}(A) = \frac{\sigma_1}{\sigma_N}$ is very large, since σ_N is very small. Due to this and m being noisy, the image x_{SVD} will be unrecognizable garbage.

This happens because $\frac{u_i^T m}{\sigma_i}$ will blow up the noise in m due to dividing by the very small σ_i for higher indices i.

To improve results, use the truncated SVD technique $x_{TSVD} = \sum_{i=1}^{T} \frac{u_i^T m}{\sigma_i} v_i$, where summation stops at an index T so the small singular values causing the issues are cut off.

b)* You try your hand now with ART, another technique to solve the linear system $Ax = m$. Explain briefly how ART works in general. What decisive advantage does ART have over SVD in this scenario?
Since the measurements m are noisy, the linear system actually has no solution. What does ART typically do in this situation?

0 1 2 3

ART is an iterative method, starting at an initial guess for x and updating it every iteration. It interprets each row of the linear system as an affine hyperplane, and repeatedly projects the current guess orthogonally to the affine hyperplanes of the linear system.

The advantage of ART is that it requires only one row of A at a time, while SVD requires the whole matrix A in memory.

If there is no solution that means the affine hyperplanes do not intersect. ART will oscillate between the affine hyperplanes endlessly, creating a so-called ‘limit cycle’.

c)* Your scan protocol of the X-ray CT scanner measures all detector positions twice, not just once. What does this mean for your system matrix A, and what consequences does it have for solving the inverse problem like in a) or b)? Explain your reasoning briefly.

0 1 2

Since each measurement corresponds to one row in A, that means every row in A has a duplicate. That means A has only $\frac{M}{2} < N$ linear independent rows and is thus now singular, i.e. it will have singular values of zero.The problem is now also underdetermined.

When solving with methods like SVD in a) this causes issues, in this particular example division by zero when dividing by the zero singular values.ART as in b) has no issues per se, even though due to detection statistics duplicate rows might have inconsistent measurements now.

Problem 5 Miscellaneous (4 credits)

Mark correct answers with a cross ☒

To undo a cross, completely fill out the answer option ■

To re-mark an option, use a human-readable marking × ■

a)* Assume the linear space $\mathbb{R}^n$ for some $n \in \mathbb{N}$. Which of these norms is induced by a scalar product?

- [ ] $\|x\|_3$
- [ ] $\|x\|_\infty$
- [x] $\|x\|_2$
- [ ] $\|x\|_1$

b)* Which statement about a linear mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ for $n, m \in \mathbb{N}$ is **not** correct?

- [ ] f is fully determined by its values $f(e_1), \ldots, f(e_n)$ for a basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$
- [ ] f can be interpreted as a vector in another linear space
- [ ] f can be represented by a matrix $A \in \mathbb{R}^{m \times n}$
- [x] f is always bijective

c)* Which condition guarantees a matrix $A \in \mathbb{R}^{n \times n}$ for $n \in \mathbb{N}$ to be invertible?

- [ ] $\operatorname{rank}(A) < n$
- [ ] $\det(A) = 0$
- [ ] $\operatorname{rank}(A) = \dim\ker(A)$
- [x] $\dim\ker(A) = 0$

d)* Which method can **not** be used to solve a least squares problem?

- [ ] pseudo inverse
- [ ] Golub's method
- [ ] normal equation
- [x] QR algorithm

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.

Computational Imaging and Inverse Problems
Department of Informatics
Technical University of Munich

Esolution
Place student sticker here

Note:
- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Basic Math Tools for Imaging and Visualization

Exam:	IN2124 / Homework-2-AN	**Date:**	Wednesday 15th December, 2021
Examiner:	PD Dr. Tobias Lasser	**Time:**	09:00 – 09:00

Working instructions

- Make sure to **sign your name** at the top of this page.
- This graded homework assignment consists of **8 pages** with a total of **6 problems**.
- The total amount of achievable credits in this graded homework assignment is 40 credits.
- Use only the pages provided in this PDF. Do not insert additional pages. There are two pages at the end with additional space for solutions. If you use them, make sure to mark which (sub)problem your answers are related to, and note in the original problem that your solution continues elsewhere.
- Allowed resources:
  - all materials, for example books, scripts, slides, sources from the internet, programs written by yourself
  - **not** allowed is help from any third person in any form
  - **not** allowed is plagiarism
- Subproblems marked by * can be solved without results of previous subproblems.
- **Answers are only accepted if the solution approach is documented.** Give a reason for each answer unless explicitly stated otherwise in the respective subproblem.
- Do not write with red or green colors nor use pencils.

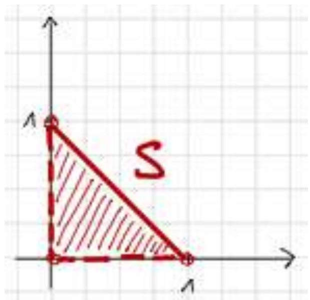
Problem 1 **Metric spaces (6 credits)**

Consider the set

$$\mathcal{S} = \left\{x = (x_1, x_2) \in \mathbb{R}^2 : \ 0 < x_1 < 1, \ 0 < x_2 < 1, \ x_1 + x_2 \le 1\right\}.$$

0 1

a)* Draw the set $\mathcal{S}$. Please make a tidy drawing, and clearly mark which points belong to the set (e.g. by using a different color, textual explanations might not hurt either).



Note: $(x, 0)$ for $x \in [0, 1]$ and $(0, y)$ for $y \in [0, 1]$ are **not** included!

0 1 2

b)* Determine whether the set $\mathcal{S}$ is open or closed in $(\mathbb{R}^2, \|\cdot\|_2)$. Use clear mathematical arguments.

$\mathcal{S}$ is not open: open balls $B(x, \varepsilon)$ with radius $\varepsilon > 0$ around the points $x = (x_1, x_2) \in \mathcal{S}$ with $x_1 + x_2 = 1$ contain both elements from $\mathcal{S}$ and from $\mathcal{S}^c$, hence $\mathcal{S}$ is not open.

$\mathcal{S}$ is not closed: $\mathcal{S}^c = \mathbb{R}^2 \setminus \mathcal{S}$ is not open, as open balls around axis points, i.e. $B\big((x_1, x_2), \varepsilon\big)$ for $(x_1, x_2) \in \mathcal{S}^c$ with $x_1 = 0$ or $x_2 = 0$, contain elements from $\mathcal{S}$ for all $\epsilon > 0$. Hence $\mathcal{S}$ is also not closed.

0 1 2 3

c)* Is the set $\mathcal{S}$ convex? Justify your answer using mathematical formulas.

$\mathcal{S}$ is convex. We need to show $\lambda x + (1 - \lambda)y \in \mathcal{S}$ for all $x, y \in \mathcal{S}$ and $\lambda \in [0, 1]$:

$x, y \in \mathcal{S} \Rightarrow x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ with $x_1, x_2, y_1, y_2 \in (0, 1)$ and $x_1 + x_2 \le 1$ and $y_1 + y_2 \le 1$

$$\Rightarrow \ z = \lambda x + (1 - \lambda)y = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix} + \begin{pmatrix} (1-\lambda)y_1 \\ (1-\lambda)y_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 + (1-\lambda)y_1 \\ \lambda x_2 + (1-\lambda)y_2 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Since $x_1, y_1 \in (0, 1)$ and $\lambda \in [0, 1]$ we have $z_1 \in (0, 1)$; same for $z_2 \in (0, 1)$.
Additionally we have

$$z_1 + z_2 = \lambda(x_1 + x_2) + (1 - \lambda)(y_1 + y_2) \le 1$$

since $x_1 + x_2 \le 1$ and $y_1 + y_2 \le 1$ and $\lambda \in [0, 1]$.
That means $z \in \mathcal{S}$ and hence $\mathcal{S}$ is convex.

Problem 2 Analysis of functions (7 credits)

a)* Calculate the Jacobi matrix of

$$f:\mathbb{R}^3\to\mathbb{R}^2,\qquad f(x,y,z)=\begin{pmatrix}6x^3y^2z+3z^4\\ z\exp(xy)\end{pmatrix}.$$

0 1 2

$$J_f(x,y,z)=\begin{pmatrix}18x^2y^2z & 12x^3yz & 6x^3y^2+12z^3\\ zy\exp(xy) & zx\exp(xy) & \exp(xy)\end{pmatrix}$$

b)* Find all local extrema of the function

$$g:\mathbb{R}^2\to\mathbb{R},\qquad g(x,y)=x^5-2x^3+x+xy^2.$$

Additionally, compute the directional derivative of the g in direction of $v=\frac{1}{\sqrt{2}}(1,1)^T$ at $(x,y)=(1,1)$.
Make sure to include all relevant steps of your computations.

0 1 2 3 4 5

$$\nabla g(x,y)=\begin{pmatrix}5x^4-6x^2+1+y^2\\ 2xy\end{pmatrix}$$

$$H_g(x,y)=\begin{pmatrix}20x^3-12x & 2y\\ 2y & 2x\end{pmatrix}$$

Candidates for local extrema require $\nabla g(x,y)=\mathbf{0}$. We have $x=0$ or $y=0$.
For $x=0$, we have $y^2+1=0$ with no real-valued solution, hence no candidates (we ignore the complex valued solutions here, as the domain of g is real-valued).
For $y=0$ we have $5x^4-6x^2+1=0$, with four candidates $x_{1,2}=\pm1$, $x_{3,4}=\pm\frac{1}{\sqrt{5}}$.

For the four candidates we have:

$$H_g(-1,0)=\begin{pmatrix}-8 & 0\\ 0 & -2\end{pmatrix}\quad H_g(1,0)=\begin{pmatrix}8 & 0\\ 0 & 2\end{pmatrix}\quad H_g(-\tfrac{1}{\sqrt{5}},0)=\begin{pmatrix}\frac{8}{\sqrt{5}} & 0\\ 0 & -\frac{2}{\sqrt{5}}\end{pmatrix}\quad H_g(\tfrac{1}{\sqrt{5}},0)=\begin{pmatrix}-\frac{8}{\sqrt{5}} & 0\\ 0 & \frac{2}{\sqrt{5}}\end{pmatrix}$$

i.e. $H_g(\pm\frac{1}{\sqrt{5}})$ is indefinite, while $H_g(-1,0)$ is negative definite and $H_g(1,0)$ is positive definite.Hence $(\pm\frac{1}{\sqrt{5}},0)$ are no local extrema of g, while $(-1,0)$ is a local maximum and $(1,0)$ is a local minimum.

For the directional derivative we use the formula $\partial_v g(x,y)=\langle\nabla g(x,y),\ v\rangle$.

$$\partial_v g(1,1)=\langle\nabla g(1,1),v\rangle=\frac{1}{\sqrt{2}}\big(1\cdot(5-6+1+1)+1\cdot2\big)=\frac{3}{\sqrt{2}}$$

Problem 3 Differentiability (9 credits)

0 1 2 3 4 5

a)* Consider the function $f : \mathbb{R} \to \mathbb{R}$ with $a, b \in \mathbb{R}$:

$$f(x) = \begin{cases} x^2 + 2, & x < 2 \\ ax + b, & x \geq 2. \end{cases}$$

Find a and b such that the function is differentiable. Justify your answer.

Everything is fine, except for $x = 2$. For f to be differentiable at point $x = 2$, it must be continuous there.

$$\lim_{x \to 2^-} f(x) = 2^2 + 2 = 6$$
$$\lim_{x \to 2^+} f(x) = 2a + b$$

Then **2a + b = 6** (1)must be satisfied for $f(x)$ to be continuous at $x = 2$.

Additionally, we must ensure the function is also differentiable at $x = 2$, that is $\lim_{x \to 2} \frac{f(x) - f(2)}{x - 2}$ exists. As $f(x)$ changes at $x = 2$, in order to compute this limit, we have to compute the two one-sided limits:

$$\lim_{x \to 2^-} \frac{f(x) - f(2)}{x - 2} = \lim_{x \to 2^-} \frac{x^2 + 2 - 6}{x - 2} = \lim_{x \to 2^-} \frac{x^2 - 4}{x - 2} \lim_{x \to 2^-} x + 2 = 4$$
$$\lim_{x \to 2^+} \frac{f(x) - f(2)}{x - 2} = \lim_{x \to 2^+} \frac{ax + b - (2a + b)}{x - 2} = \lim_{x \to 2^+} \frac{a(x - 2)}{x - 2} = \lim_{x \to 2^+} a = a$$

Hence, we must have **a = 4** (2).

From (1) and (2) we conclude $a = 4$ and $b = -2$for the the function $f(x)$ to be differentiable at $x = 2$.

0 1 2 3 4

b)* Consider the function $g : \Omega \to \mathbb{R}$ with $g(x) = \max(0, x)$. Is g differentiable for $\Omega = \mathbb{R}$? Is g differentiable for $\Omega = [0, \infty)$? Compute the subdifferential of g for $\Omega = \mathbb{R}$. Justify your answers.

- $\Omega = \mathbb{R}$: g is not differentiable, as we have $\lim_{x \to 0^+} g'(x) = 1$ in contradiction to $\lim_{x \to 0^-} g'(x) = 0$, i.e. g is not differentiable in 0
- $\Omega = [0, \infty)$: g is now differentiable with $g'(x) = 1$ everywhere, including at 0

The subdifferential of g for $\Omega = \mathbb{R}$ is:

- for $x < 0$: $\partial g(x) = \{0\}$, as g is convex and differentiable for $x < 0$
- for $x > 0$: $\partial g(x) = \{1\}$, as g is convex and differentiable for $x > 0$
- for $x = 0$: $\partial g(0) = [0, 1]$, as we have $g(z) \geq cz$ for $c \in [0, 1]$.

Problem 4 More derivatives (8 credits)

a)* Consider the two-dimensional rotation matrix

$$R(\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}$$

for $\phi \in (-\pi, \pi]$. Derive an approximation of $R(\phi)$ which does not contain cos and sin and which is accurate for small angles ϕ, i.e., angles close to zero. Justify your answer using mathematical tools from chapter 2 of BMT.

0 1 2 3 4

We consider the Taylor expansions of cos,sin around 0:

$$\cos(0+\varepsilon) = \cos(0) - \sin(0)\varepsilon + rest(\varepsilon^2),$$

$$\sin(0+\varepsilon) = \sin(0) + \cos(0)\varepsilon + rest(\varepsilon^2).$$

As $\sin(0) = 0$ and $\cos(0) = 1$ we find that $\cos(0+\varepsilon) \approx 1$ and $\sin(0+\varepsilon) \approx \varepsilon$. Thus, we have that

$$R(\phi) \approx \begin{pmatrix} 1 & -\phi \\ \phi & 1 \end{pmatrix}$$

for ϕ small.

b)* Consider an image $I : \Omega \to \mathbb{R}$ with $\Omega \in \mathbb{R}^2$ open. Assume that you know the values of I only at a discrete point grid $G := \{(x_i, y_j) : i, j \in \{0, \dots, n\}\} \subset \Omega$ with spacing 2 for some $n \in \mathbb{N}$. Describe briefly how you can compute an approximation to the image gradient of I at every point of G. Justify your answer using mathematical tools from chapter 2 of BMT.

0 1 2 3 4

Just computing the gradient as usual does not work, as for the partial derivatives you need to evaluate I outside those grid points.
Instead, we do a Taylor expansion in both arguments for $(x, y) \in \Omega$ and some $h \in \mathbb{R}$:

$$I(x+h, y) \approx I(x, y) + \partial_1 I(x, y)h \quad \Rightarrow \quad \partial_1 I(x, y) \approx \frac{I(x+h, y) - I(x, y)}{h}$$

$$I(x, y+h) \approx I(x, y) + \partial_2 I(x, y)h \quad \Rightarrow \quad \partial_2 I(x, y) \approx \frac{I(x, y+h) - I(x, y)}{h}$$

Since our grid spacing is 2, we can now set $h = 2$, so that in the previous formulas we only use the discrete grid points (provided we start at an appropriate $(x, y) \in G$) to compute the approximation to gradient of I using the approximated partial derivatives.At the borders of G, we have to use some tricks like zero-padding or mirroring, so that we can still evaluate the formulas for the partial derivatives.

Problem 5 Optimization (6 credits)

Consider the following optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x), \qquad \text{where} \quad f(x) = \frac{1}{2}x^T Ax - b^T x$$

with $A \in \mathbb{R}^{n \times n}$ symmetric positive definite and $b \in \mathbb{R}^n$ with $n \in \mathbb{N}$.

0 1 2

a)* What is necessary for this optimization problem to have a solution? Explain your reasoning.

We have $f(x) : \mathbb{R}^n \to \mathbb{R}$ with $\nabla f(x) = Ax - b$.
The necessary condition to have a solution is $\nabla f(x) = 0$, i.e. $Ax - b = 0$.

Since A is symmetric positive definite, it is invertible and the linear system $Ax = b$ has a (unique) solution. Thus there is a candidate for a minimizer for f, and the necessary condition for this optimization problem to have a solution is fulfilled.

0 1 2

b) Now we want a definite answer: is there a solution to this optimization problem, and if so, is it unique? Explain your reasoning.

We reuse the necessary condition from a), i.e. that for the x^* solving $Ax = b$ we have $\nabla f(x^*) = 0$.

We have $H_f(x) = A$ and since $A > 0$, we know that f has a local minimum at x^*.

Furthermore, since $H_f(x) = A > 0$ for all $x \in \mathbb{R}^n$, we know that f is strictly convex and the minimizer x^* is unique.

0 1 2

c)* In our Case Study in chapter 2 of BMT (Lucas Kanade affine template tracking) we try to track a point in a video sequence using the model $I_s(x + v) = I_t(x)$, where $I_s, I_t : \Omega \to \mathbb{R}$ for some $\Omega \subset \mathbb{R}^2$ open. Which two key steps are required to transform this problem to a linear least squares problem? Explain briefly what assumptions underlie those key steps.

- In order to recover both coordinates of $v = (v_1, v_2) \in \mathbb{R}^2$, we have to look at more than one point $x \in \Omega$, e.g. at an entire patch $P \subset \Omega$ that contains x. This requires the assumption that v is constant over the entire patch.
- To reduce the resulting non-linear optimization problem $\min_{v \in \mathbb{R}^2} \int_P |I_s(x + v) - I_t(x)|^2 \, dx$ to a linear least squares problem, we use a Taylor expansion to linearize the problem. This requires I_s to be differentiable.(Although this requirement can be kind of ignored through use of finite differences in the discrete formulation.)

Problem 6 Miscellaneous (4 credits)

Mark correct answers with a cross ☒

To undo a cross, completely fill out the answer option ■

To re-mark an option, use a human-readable marking ×■

a)* Let $f : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ open, for some $n, m \in \mathbb{N}$. Which statement is *incorrect*?

- ☐ if f is differentiable, then f is partially differentiable
- ☒ if f is partially differentiable, then f is differentiable
- ☐ if f is continuously differentiable, then f is differentiable

b)* Which of these linear spaces is a Hilbert space? (for some $n \in \mathbb{N}$)

- ☐ $\mathbb{R}^n$ with $\|\cdot\|_\infty$
- ☒ $\mathbb{R}^n$ with $\|\cdot\|_2$
- ☐ $\mathbb{R}^n$ with $\|\cdot\|_1$

c)* Let (M, d) be a metric space and $U, V \subset M$ open sets. Which of these sets is closed?

- ☐ $\mathrm{int}(U)$
- ☐ $U \cap V$
- ☐ $U \cup V$
- ☒ ∂V

d)* Which mathematical concept is usually *not* involved when solving a minimization problem?

- ☐ Gradient
- ☐ Hessian
- ☐ Gateaux derivative
- ☒ Laplacian

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.

Computational Imaging and Inverse Problems
Department of Informatics
Technical University of Munich

Esolution
Place student sticker here

Note:
- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Basic Math Tools for Imaging and Visualization

Exam:	IN2124 / Homework-3-OPT	**Date:**	Wednesday 12th January, 2022
Examiner:	PD Dr. Tobias Lasser	**Time:**	09:00 – 09:00

Working instructions

- Make sure to **sign your name** at the top of this page.
- This graded homework assignment consists of **6 pages** with a total of **4 problems**.
- The total amount of achievable credits in this graded homework assignment is 30 credits.
- Use only the pages provided in this PDF. Do not insert additional pages. There are two pages at the end with additional space for solutions. If you use them, make sure to mark which (sub)problem your answers are related to, and note in the original problem that your solution continues elsewhere.
- Allowed resources:
  - all materials, for example books, scripts, slides, sources from the internet, programs written by yourself
  - **not** allowed is help from any third person in any form
  - **not** allowed is plagiarism
- Subproblems marked by * can be solved without results of previous subproblems.
- **Answers are only accepted if the solution approach is documented.** Give a reason for each answer unless explicitly stated otherwise in the respective subproblem.
- Do not write with red or green colors nor use pencils.

Problem 1 Least squares problems (9 credits)

Consider the least squares problem

$$\min_{x \in \mathbb{R}^n} f(x), \qquad f(x) = \tfrac{1}{2}\|Ax - b\|_2^2,$$

where $A \in \mathbb{R}^{m \times n}$ is full rank, and $b \in \mathbb{R}^m$, for $n, m \in \mathbb{N}$ and $m < n$.

0 1 2

a)* Compute the gradient and the Hessian of f.

We have the gradient

$$\nabla f(x) = A^T(Ax - b) = A^TAx - A^Tb.$$

and the Hessian

$$H_f(x) = A^TA.$$

0 1 2 3 4

b) Assume now $A = \begin{pmatrix} 1 & -1 \end{pmatrix} \in \mathbb{R}^{1\text{x}2}$. Is the *sufficient* condition for our least squares problem to have a solution fulfilled for this particular A? Explain your reasoning mathematically.

First, we need $\nabla f(x) = 0$, i.e. $A^TAx = A^Tb$ following a). We have $A^TA = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $A^Tb = \begin{pmatrix} b \\ -b \end{pmatrix}$.

Hence for $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ we need

$$A^TAx = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ -x_1 + x_2 \end{pmatrix} \overset{!}{=} \begin{pmatrix} b \\ -b \end{pmatrix} = A^Tb,$$

which has infinitely many solutions.

Second, we need $H_f(x) = A^TA$ to be positive definite. However, A^TA here has $\det(A^TA) = 0$, hence has 0 as eigenvalue and thus is not positive definite.
The sufficient condition is thus not fulfilled.

0 1 2 3

c) Now we want to solve our least squares problem using gradient descent. Formulate a suitable step size α_k for the k-th step of gradient descent, and explain your reasoning mathematically.

Similar to the gradient descent for quadratic forms, we want to optimize the step size α_k by solving

$$\frac{\partial}{\partial \alpha_k} f(x_k + \alpha_k d_k) \overset{!}{=} 0$$

for α_k, where $d_k = -\nabla f(x)$.
We have

$$\begin{aligned} \frac{\partial}{\partial \alpha_k} f(x_k + \alpha_k d_k) &= \nabla f(x_k + \alpha_k d_k)^T d_k = \left(A^TA(x_k + \alpha_k d_k) - A^Tb\right)^T d_k \\ &= (A^TAx_k - A^Tb)^T d_k + \alpha_k (A^TAd_k)^T d_k = -d_k^T d_k + \alpha_k d_k^T A^TAd_k \end{aligned}$$

and hence

$$\alpha_k = \frac{d_k^T d_k}{d_k^T A^TAd_k}$$

Problem 2 Gradient-based methods (7 credits)

Let us now consider a regularized optimization problem,

$$\min_{x \in \mathbb{R}^n} D(x, y) + \lambda R(x)$$

with $D : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $R : \mathbb{R}^n \to \mathbb{R}$ both continuously differentiable, a regularization parameter $\lambda > 0$, some constant $y \in \mathbb{R}^m$, and $n, m \in \mathbb{N}$.

a)* You want to attempt to use the gradient descent method to solve our optimization problem. Describe two different strategies to choose the step size. Explain briefly the advantages and short-comings of each of them.

0 1 2 3 4

The simplest strategy is to choose a small, fixed step size.
Advantage: simple to implement. Disadvantage: hard to choose (if it's too big, it will not converge at all, if it's too small, it will have very low convergence speed).

Another strategy is to use a line-search algorithm, i.e. trying to optimize the step size for each iteration separately.
Advantage: will yield (near) optimal steps, i.e. fast convergence speed. Disadvantage: hard to implement, and more computational effort (additional optimization in each iteration).

Third alternative: decaying step sizes. Advantage: simple to implement, and can have relatively big steps at starts (to get closer to solution fast), and smaller and smaller steps later (to make sure you actually reach the solution). Disadvantage: even harder to choose (first the initial value, then the rate of decay, if it's too big, it will not converge at all, if it's too small, it will have very low convergence speed).

b)* Formulate the subgradient method for our minimization problem as a fixed point iteration of the function $T : \mathbb{R}^n \to \mathbb{R}^n$. Specify T and ensure that all terms are well-defined.

0 1

Fixed point iteration $x_{k+1} = T(x_k)$ using $T(x_k) := x_k - \alpha_k g_k$, where α_k is the step size at step $k \in \mathbb{N}$, and g_k is any subgradient of $f(x) = D(x, y) + \lambda R(x)$ at x_k (assuming it exists).(Since f is actually differentiable, if it were convex we could choose $g_k = \nabla f(x_k)$.)

c)* Let us assume now that $D(x, y) = \frac{1}{2}\|Ax - y\|_2^2$ for $A \in \mathbb{R}^{m \times n}$ with full rank for $n, m \in \mathbb{N}$ with $m > n$, and that $R(x) = 0$. You try using the conjugate gradient method for this particular optimization problem, but you fail. Explain briefly why you fail. State a suitable fast method with similar properties from BMT chapter 3 to solve your problem, and explain briefly why it is applicable.

0 1 2

A is not square and not symmetric positive definite, hence CG cannot be applied.
However, by using CG on the normal equation (CGNE), i.e. solving $A^T Ax = A^T y$ using CG, the solution to our optimization problem can be computed. CG is applicable on the normal equation, as $A^T A$ is square and symmetric positive definite.

Problem 3 Dealing with noise (8 credits)

Consider the noisy linear system

$$Ax + \epsilon = y,$$

where $A \in \mathbb{R}^{n\times n}$ is an ill-conditioned invertible matrix, $\epsilon \in \mathbb{R}^n$ denotes unknown noise, and $y \in \mathbb{R}^n$ some right hand side for $n \in \mathbb{N}$.

0 1 2 3

a)* We try to solve the linear system using A^{-1}, and we get $x + A^{-1}\epsilon = A^{-1}y$. Explain mathematically why the noise term $A^{-1}\epsilon$ can be a huge problem, even if the noise is very small, i.e. if $\|\epsilon\|_2$ is very small. Use coherent mathematical arguments and be brief.
You may use the induced matrix norm $\|A\|_2 = \sigma_1$, where $\sigma_1 \geq \sigma_2 \geq ... \sigma_n$ are the singular values of A, and the following inequality: $\|Ax\|_2 \leq \|A\|_2\|x\|_2$.

We know that A is ill-conditioned, i.e. cond$(A) = \frac{\sigma_1}{\sigma_n}$ is large. That implies that σ_n is very small.

We know that $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$, i.e. very large, as A^{-1} has singular values $\frac{1}{\sigma_n} \geq ... \geq \frac{1}{\sigma_1}$.

Using the given inequality, we have $\|A^{-1}\epsilon\|_2 \leq \|A^{-1}\|_2\|\epsilon\|_2$. So even if $\|\epsilon\|_2$ is very small, $\|A^{-1}\|_2$ can be very large, and thus can cause a huge problem when solving the noisy linear system.

0 1 2

b)* There are two methods in the BMT toolbox that help in dealing with such noisy linear systems. State the names of the two methods. Even though both methods are quite different, they both rely on the same mathematical principle to help in dealing with such noisy linear systems. Explain that principle briefly.

The two methods are: truncated SVD and Tikhonov regularization.
The shared mathematical principle is that both methods are dealing with the very small singular values in ill-conditioned matrices (which are causing the issues). Truncated SVD cuts them off, while Tikhonov regularization weights them down.

0 1 2 3

c)* We now want to compute a Tikhonov regularized solution, i.e. we compute $T_\lambda(y) = \arg\min_{z\in\mathbb{R}^n}\|Az - y\|_2^2 + \lambda\|z\|_2^2$ for some $\lambda > 0$. Assume now that n is very large. Briefly describe a practical approach to compute the concrete solution $T_\lambda(y)$.

Since the matrix A is very large, using SVD is impractical. Hence we use the stacked form,i.e. we use two linear systems, $Az = y$ and $\sqrt{\lambda}Iz = 0$, stacked on top of each other.
The least squares solution of this is computed via the normal equation, yielding $(A^TA + \lambda I)z = A^Ty$.
We solve the normal equation for z using CG (conjugate gradient) to receive $T_\lambda(y)$, as the matrix $A^TA + \lambda I$ is symmetric positive definite (because A is invertible and $\lambda > 0$).

Problem 4 Miscellaneous (6 credits)

Mark correct answers with a cross ☒
To undo a cross, completely fill out the answer option ■
To re-mark an option, use a human-readable marking ×■

a)* You perform the conjugate gradient method. Consider the last two directions of improvement d_{n-1} and d_n. What is the angle between d_{n-1} and d_n regarding the energy scalar product?

- ☐ 180°
- ☒ 90°
- ☐ 0°

b)* Consider the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$. What is the subdifferential of f at 0?

- ☐ $[0, 1]$
- ☐ $[-1, 1]$
- ☒ $\{1\}$

c)* You apply the Newton method to a non-linear problem. Which statement is correct?

- ☒ It does not converge, as your initial value was too far away from the solution.
- ☐ The Newton method cannot be applied to a non-linear problem.
- ☐ It converges, as it works for all initial values.

d)* What is a decent method for choosing the regularization parameter in Tikhonov regularization?

- ☐ Golub's method
- ☐ Brute-force method
- ☒ L-curve method

e)* You do a fixed point iteration for $T : [0, 1] \to [0, 1]$. What happens?

- ☒ It converges if T is a contraction.
- ☐ It always converges.
- ☐ It does not converge if T is a contraction.

f)* Why would you consider using preconditioning for your optimization problem?

- ☐ To find a unique solution.
- ☐ To increase the condition of the matrix involved.
- ☒ To improve convergence speed of your solver.

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.

Computational Imaging and Inverse Problems
Department of Informatics
Technical University of Munich

Esolution
Place student sticker here

Note:
- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Basic Math Tools for Imaging and Visualization

Exam:	IN2124 / Homework-4-PT	**Date:**	Wednesday 2nd February, 2022
Examiner:	PD Dr. Tobias Lasser	**Time:**	09:00 – 09:00

Working instructions

- Make sure to **sign your name** at the top of this page.
- This graded homework assignment consists of **6 pages** with a total of **3 problems**.
- The total amount of achievable credits in this graded homework assignment is 20 credits.
- Use only the pages provided in this PDF. Do not insert additional pages. There are two pages at the end with additional space for solutions. If you use them, make sure to mark which (sub)problem your answers are related to, and note in the original problem that your solution continues elsewhere.
- Allowed resources:
  - all materials, for example books, scripts, slides, sources from the internet, programs written by yourself
  - **not** allowed is help from any third person in any form
  - **not** allowed is plagiarism
- Subproblems marked by * can be solved without results of previous subproblems.
- **Answers are only accepted if the solution approach is documented.** Give a reason for each answer unless explicitly stated otherwise in the respective subproblem.
- Do not write with red or green colors nor use pencils.

Problem 1 Probability spaces and random variables (6 credits)

Let us assume that we have a coin that is deformed and weighted. When you toss the coin, it lands on its side in 80% of the tosses. The coin lands on heads 15% of the time, and otherwise on tails.

0 1 2 3

a)* Explicitly write down a suitable probability space for a single coin toss experiment using this deformed and weighted coin. Be sure to be explicit, i.e. write out everything, and be mathematically precise.

probability space $(\Omega, \mathcal{F}, P)$ with:

- sample space: $\Omega = \{S, H, T\}$ (side, heads, tails)
- event space: $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{S\}, \{H\}, \{T\}, \{S, H\}, \{S, T\}, \{H, T\}, \Omega\}$
- probability measure $P : \mathcal{F} \to \mathbb{R}$,
  - $P(\emptyset) = 0$
  - $P(\Omega) = 1$
  - $P(\{S\}) = 0.8$
  - $P(\{H\}) = 0.15$
  - $P(\{T\}) = 0.05$
  - $P(\{S, H\}) = 0.95$
  - $P(\{S, T\}) = 0.85$
  - $P(\{H, T\}) = 0.2$

0 1

b)* We are now tossing our deformed, weighted coin several times until it lands on its side. Let us denote Y as the number of times the coin was actually tossed. Is Y a random variable? Explain your reasoning.

Yes, as every function on a countably infinite sample space is measurable, and a sample space of tossing a coin repeatedly is countably infinite.(note: the sample space, however you define it, has to be countably infinite, as you do not know how often you have to toss the coin.)

0 1 2

c) Assuming now that $Y = 13$ (where Y is from 1b), what is the expected number of tails tossed before the coin landed on its side? Use the conditional expectation and explain your result.

We use the random variable Y from the previous assignment with $Y = 13$.
We denote the number of tails tossed as a random variable X, which has a Binomial distribution, $X \sim B(n, p)$, with parameters $n = 13 - 1 = 12$ and $p = \frac{5}{15+5} = 0.25$.
The expected number of tails is thus the conditional expectation, given the condition that $Y = 13$, i.e.

$$E(X|Y = 13) = np = 12 \cdot 0.25 = 3.$$

Problem 2 Estimators (8 credits)

Let us assume that we have data points $(z_i)_{i=1}^n$ with $z_i \in \mathbb{R}$ for $i = 1, \dots, n$ and $n \in \mathbb{N}$. We further assume that our data is independently and identically distributed according to a normal distribution with unknown parameters $\theta = (\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

a)* Write down the probability density function for a single data point z_i.

0 1

The z_i are sampled from a normally distributed random variable, thus

$$f(z_i; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z_i - \mu)^2}{2\sigma^2}\right)$$

b) Write down the likelihood function for all data points $(z_i)_{i=1}^n$. Explain your reasoning.

0 1 2

The likelihood function is

$$l(z_1, \dots, z_n; \theta) = \prod_{i=1}^n f(z_i; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z_i - \mu)^2}{2\sigma^2}\right)$$

It is the product of the probability density functions $f(z_i; \theta)$ because the data points are independentlyand identically distributed (iid).

c) Derive a maximum likelihood estimator for the parameter μ. Explain your result briefly in words. Make sure to list all important steps of your derivation, and justify all of your computations mathematically.

0 1 2 3 4 5

We compute a maximum likelihood estimator, i.e.

$$\mu_{ML} = \arg\max_\mu l(z_1, \dots, z_n; \theta) = \prod_{i=1}^n f(z_i; \theta).$$

To simplify computations, we maximize the log-likelihood instead, which preserves the location of the maximum and eases computations:

$$\begin{aligned}
\mu_{ML} &= \arg\max_\mu L(z_1, \dots, y_n; \theta) = \sum_{i=1}^n \ln f(z_i; \theta) \\
&= \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z_i - \mu)^2}{2\sigma^2}\right)\right) \\
&= -\tfrac{n}{2} \ln(2\pi) - n \ln(\sigma) - \tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (z_i - \mu)^2
\end{aligned}$$

To find an estimate for μ, we differentiate the log-likelihood with respect to μ, set it to zero and solve for μ:

$$\frac{d}{d\mu} L(z_1, \dots, z_n; \theta) = \frac{1}{\sigma^2} \sum_{i=1}^n (z_i - \mu) \overset{!}{=} 0,$$

hence we have

$$\frac{1}{\sigma^2} \sum_{i=1}^n z_i = \frac{n}{\sigma^2} \mu \quad \text{and thus} \quad \mu = \frac{1}{n} \sum_{i=1}^n z_i.$$

In words: the maximum likelihood estimator of μ is the mean of the data(note: regardless of σ^2).

Problem 3 Miscellaneous (6 credits)

Mark correct answers with a cross ☒
To undo a cross, completely fill out the answer option ■
To re-mark an option, use a human-readable marking × ■

a)* What is the Borel-σ-field of $\mathbb{R}$?

- ☐ It is the set of all open sets in $\mathbb{R}$.
- ☐ It is the power set of $\mathbb{R}$.
- ☒ It is the σ-field created by the open sets of $\mathbb{R}$.

b)* Let X be a continuous random variable. How can you uniquely define the distribution of X?

- ☐ With the covariance matrix.
- ☐ With the probability mass function.
- ☒ With the probability density function.

c)* Let X, Y be two uncorrelated random variables. Which statement is true?

- ☐ X, Y are independent.
- ☐ $E(X|Y) = E(X)$.
- ☒ If (X, Y) Gaussian, then X, Y are independent.

d)* Consider an estimator T of parameters θ. What is its bias?

- ☒ It is the error of the expected estimator, $E(T) - \theta$.
- ☐ It is the prejudice of the estimator, $Pr(T)$.
- ☐ It is the error of the estimator, $T - \theta$.

e)* What does the expectation maximization method guarantee?

- ☐ That you maximize the expectation.
- ☐ There are no guarantees.
- ☒ That you do not decrease the expectation.

f)* What is the MLEM algorithm used for in clinical practice?

- ☐ MRI
- ☒ SPECT
- ☐ X-ray CT

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.